

# CALDERÓN COMMUTATORS AND THE CAUCHY INTEGRAL ON LIPSCHITZ CURVES REVISITED I. FIRST COMMUTATOR AND GENERALIZATIONS

CAMIL MUSCALU

**ABSTRACT.** This article is the first in a series of three papers, whose scope is to give new proofs to the well known theorems of Calderón, Coifman, McIntosh and Meyer [1], [8], [9]. Here we treat the case of the first commutator of Calderón and some of its generalizations.

## 1. INTRODUCTION

This is the first paper in a sequel of three, whose aim is to give new proofs to the well known theorems of Calderón, Coifman, McIntosh and Meyer [1], [8], [9], which established  $L^p$  estimates for the so called Calderón commutators and the Cauchy integral on Lipschitz curves.

We refer the reader to the book of Coifman and Meyer [8] for a description of the history of these fundamental analytical objects, the role they play in analysis and the various methods that have been further developed to understand these operators, since the appearance of the original articles.

Other expository papers where some of these results are described and connected with other parts of mathematics, are the proceedings of the plenary talks at the ICM 1974 in Vancouver and the ICM 1978 in Helsinki, given by Fefferman [12] and Calderón [2].

Our approach will also turn out to be flexible and generic enough, to allow us to generalize these classical results in various new ways.

The first paper describes the case of the first commutator and its generalizations, the second one the case of the Cauchy integral on Lipschitz curves and its generalizations and finally, the third will be devoted to the extension of all these results to the multi-parameter setting of polydiscs of arbitrary dimension, solving completely along the way an open question of Coifman from the early eighties.

We naturally start with the first commutator.

Given a Lipschitz function  $A$  on the real line (so  $A' := a \in L^\infty(\mathbf{R})$ ) one formally defines the linear operator  $C_1(f)$  by the formula

$$C_1(f)(x) = p.v. \int_{\mathbf{R}} \frac{A(x) - A(y)}{(x - y)^2} f(y) dy \quad (1)$$

where the meaning of the principal value integral is

$$\lim_{\epsilon \rightarrow 0} \int_{\epsilon < |x-y| < 1/\epsilon} \frac{A(x) - A(y)}{(x - y)^2} f(y) dy \quad (2)$$

whenever the limit exists. This is the so called first commutator of Calderón. Note that the simplest particular case is obtained when  $A(x) = x$  and  $C_1(f)$  becomes the classical Hilbert transform.

Let us observe that when  $a$  and  $f$  are Schwartz functions, then (2) makes perfect sense.

Indeed, for a fixed  $\epsilon > 0$ , one can rewrite the corresponding expression in (2) as

$$\begin{aligned} - \int_{\epsilon < |t| < 1/\epsilon} \frac{A(x+t) - A(x)}{t^2} f(x+t) dt &= - \int_{\epsilon < |t| < 1/\epsilon} \left[ \frac{A(x+t) - A(x)}{t} \right] f(x+t) \frac{dt}{t} = \\ &= - \int_{\epsilon < |t| < 1/\epsilon} \left[ \int_0^1 a(x+\alpha t) d\alpha \right] f(x+t) \frac{dt}{t}. \end{aligned} \quad (3)$$

Then, write  $a$  and  $f$  as

$$a(x+\alpha t) = \int_{\mathbf{R}} \widehat{a}(\xi_1) e^{2\pi i(x+\alpha t)\xi_1} d\xi_1$$

and

$$f(x+t) = \int_{\mathbf{R}} \widehat{f}(\xi) e^{2\pi i(x+t)\xi} d\xi.$$

Using these formulas in (3), the expression becomes

$$- \int_{\mathbf{R}^2} m_\epsilon(\xi, \xi_1) \widehat{f}(\xi) \widehat{a}(\xi_1) e^{2\pi i x(\xi + \xi_1)} d\xi d\xi_1 \quad (4)$$

where

$$m_\epsilon(\xi, \xi_1) = \int_0^1 \int_{\epsilon < |t| < 1/\epsilon} \frac{1}{t} e^{2\pi i t(\xi + \alpha \xi_1)} dt d\alpha$$

which is known to converge uniformly to

$$- \int_0^1 \operatorname{sgn}(\xi + \alpha \xi_1) d\alpha.$$

In particular, the dominated convergence theorem implies that the limit as  $\epsilon \rightarrow 0$  exists in (4) and it is equal to

$$\int_{\mathbf{R}^2} \left[ \int_0^1 \operatorname{sgn}(\xi + \alpha \xi_1) d\alpha \right] \widehat{f}(\xi) \widehat{a}(\xi_1) e^{2\pi i x(\xi + \xi_1)} d\xi d\xi_1. \quad (5)$$

Because of this formula, one can think of  $C_1$  as being a bilinear operator in  $f$  and  $a$  and we will denote it from now on with  $C_1(f, a)$ . The following theorem of Calderón is classical [1].

**Theorem 1.1.** *For every  $A' = a \in L^\infty$ , the operator  $C_1$  extends naturally as a bounded linear operator from  $L^p(\mathbf{R})$  into  $L^p(\mathbf{R})$  for every  $1 < p < \infty$ , satisfying*

$$\|C_1(f, a)\|_p \lesssim \|a\|_\infty \cdot \|f\|_p. \quad (6)$$

At this point, one should observe that the symbol of (5) given by

$$(\xi, \xi_1) \rightarrow \int_0^1 \operatorname{sgn}(\xi + \alpha \xi_1) d\alpha \quad (7)$$

is not a Marcinkiewicz-Hörmander-Mihlin symbol [25] and as a consequence, the Coifman-Meyer theorem on paraproducts [8] cannot be applied. More precisely, one can see that away from the lines  $\xi = 0$  and  $\xi + \xi_1 = 0$ , the symbol (7) is many times differentiable and behaves like a classical symbol, but along them it is only continuous. The observation on which our approach is based, is that in spite of this lack of differentiability, when one *smoothly* restricts (7) to an arbitrary Whitney square with respect to the origin <sup>1</sup>, the Fourier coefficients of the corresponding function decay at least quadratically. This fact (which will be proved carefully later on) will reduce the problem to the one of proving estimates for the associated bilinear operators, which do not grow too fast with respect to the indices of the Fourier coefficients. We will see that this upper bounds can grow at most logarithmically, which will be more than enough to make the final power series convergent. This is, in just a few words, our strategy of the proof.

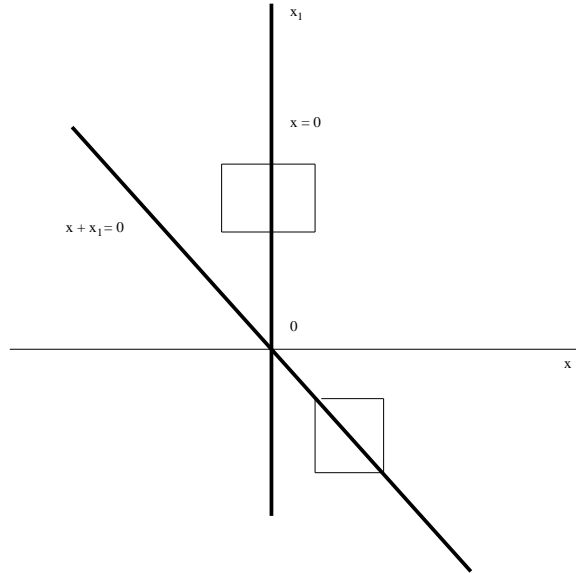


FIGURE 1. The singularities of the symbol of the first commutator

Before going any further, let us also remark that if one permutes the two integrations in (5), one can rewrite that expression as

$$\int_0^1 BHT_\alpha(f, a)(x) d\alpha$$

where  $BHT_\alpha$  is the so called bilinear Hilbert transform of parameter  $\alpha$ . An alternative approach to the first commutator (suggested by Calderón), was to prove  $L^p \times L^\infty \rightarrow L^p$  estimates for these operators, with implicit constants that are integrable or even uniform in  $\alpha$ . Estimates for the bilinear Hilbert transform have been first proved by Lacey and Thiele in [15] and [16], and uniform estimates have been later on obtained by Thiele [26], Grafakos and Li [13] and Li [17]. It is also interesting to remark that it is not yet known whether such an approach would work for the second Calderón commutator <sup>2</sup> which this time can be written as

<sup>1</sup>These are squares whose sides are parallel to the coordinate axes and whose distance to the origin is comparable to their sidelengths.

<sup>2</sup>The second commutator can similarly be seen as a trilinear operator with symbol  $\int_{[0,1]^2} \text{sgn}(\xi + \alpha\xi_1 + \beta\xi_2) d\alpha d\beta$ .

$$\int_{[0,1]^2} THT_{\alpha\beta}(f, a, a)(x) d\alpha d\beta.$$

A recent result of Palsson [24] proved many estimates for the operator  $\int_0^1 THT_{\alpha\beta} d\alpha$  ( $\beta$  is fixed now), but so far there are no  $L^p$  estimates available (uniform or not) for the corresponding trilinear operator  $THT_{\alpha\beta}$  which has been named by several authors, the trilinear Hilbert transform.

Now coming back to (6), in order to describe the way in which  $C_1(f, a)$  can be defined for any  $a \in L^\infty$  and any  $f \in L^p(\mathbf{R})$ , we need to say a few words about adjoints of bilinear operators.

If  $m(\xi_1, \xi_2)$  is a bounded symbol, denote by  $T_m(f_1, f_2)$  the bilinear operator given by

$$T_m(f_1, f_2)(x) = \int_{\mathbf{R}^2} m(\xi_1, \xi_2) \widehat{f_1}(\xi_1) \widehat{f_2}(\xi_2) e^{2\pi i x(\xi_1 + \xi_2)} d\xi_1 d\xi_2, \quad (8)$$

for  $f_1, f_2$  Schwartz functions. Associated with it is the trilinear form  $\Lambda(f_1, f_2, f_3)$  defined by

$$\Lambda(f_1, f_2, f_3) = \int_{\mathbf{R}} T_m(f_1, f_2)(x) f_3(x) dx,$$

again for  $f_1, f_2, f_3$  Schwartz functions.

There are two adjoint operators  $T_m^{*1}$  and  $T_m^{*2}$  naturally defined by the equalities

$$\int_{\mathbf{R}} T_m^{*1}(f_2, f_3)(x) f_1(x) dx = \Lambda(f_1, f_2, f_3)$$

and

$$\int_{\mathbf{R}} T_m^{*2}(f_1, f_3)(x) f_2(x) dx = \Lambda(f_1, f_2, f_3)$$

respectively. It is very easy to observe that both of them are also bilinear multipliers whose symbols are  $m(-\xi_1 - \xi_2, \xi_2)$  and  $m(\xi_1, -\xi_1 - \xi_2)$  respectively.

Now, if  $a$  and  $f$  are Schwartz functions, the inequality (6) is equivalent to

$$\left| \int_{\mathbf{R}} C_1(f, a)(x) g(x) dx \right| \lesssim \|a\|_\infty \cdot \|f\|_p \cdot \|g\|_{p'} \quad (9)$$

for any Schwartz function  $g$ , where  $p'$  is the dual index of  $p$  (so  $1/p + 1/p' = 1$ ). We also know from the above that

$$\int_{\mathbf{R}} C_1(f, a)(x) g(x) dx = \int_{\mathbf{R}} C_1^{*2}(f, g)(x) a(x) dx. \quad (10)$$

We are going to prove in the rest of the paper that

$$\|C_1^{*2}(f, g)\|_1 \lesssim \|f\|_p \cdot \|g\|_{p'} \quad (11)$$

for any Schwartz functions  $f, g$  and this shows that  $C_1^{*2}$  can be extended by density to the whole  $L^p \times L^{p'}$ . But then, this means that the right hand side of (10) makes sense for any  $a \in L^\infty$ , not only for bounded Schwartz functions, and this suggests to extend  $C_1(f, a)$  by duality. More specifically, for  $f \in L^p$  and  $a \in L^\infty$ , one can define  $C_1(f, a)$  to be the unique  $L^p$  function satisfying (10) for any  $g \in L^{p'}$ .

This discussion also proves that to demonstrate Theorem 1.1, we only need to prove (11). The idea now is to discretize  $C_1^{*2}$  and reduce (11) to a discrete finite model.

**Acknowledgement:** We wish to thank Eli Stein who after a talk we gave in Pisa, kindly pointed out to us that the maximal Theorem 4.1 that will enter the picture later on, was actually known and can be found in [25] Chapter II.

The present work has been partially supported by the NSF.

## 2. REDUCTION TO A FINITE LOCALIZED MODEL

We start with some standard notations and definitions. An interval  $I$  on the real line  $\mathbf{R}$  is called dyadic if it is of the form  $I = [2^k n, 2^k(n+1)]$  for some  $k, n \in \mathbf{Z}$ . We will denote by  $\mathcal{D}$  the set of all such dyadic intervals.

If  $I \in \mathcal{D}$ , we say that a smooth function  $\Phi_I$  is a bump adapted to  $I$  if and only if one has

$$|\partial^\alpha(\Phi_I)(x)| \leq C_{\alpha,N} \cdot \frac{1}{|I|^\alpha} \cdot \frac{1}{\left(1 + \frac{\text{dist}(x,I)}{|I|}\right)^N}$$

for every integer  $N$  and sufficiently many derivatives  $\alpha$ , where  $|I|$  is the length of  $I$ . If  $\Phi_I$  is a bump adapted to  $I$ , we say that  $|I|^{-1/p}\Phi_I$  is an  $L^p$  - normalized bump adapted to  $I$ , for  $1 \leq p \leq \infty$ . Also, if  $I \in \mathcal{D}$  and  $\mathbf{n} \in \mathbf{Z}$  we denote by  $I_{\mathbf{n}}$  the new dyadic interval  $[2^k(n - \mathbf{n}), 2^k(n + 1 - \mathbf{n})]$  sitting  $\mathbf{n}$  units of length  $|I|$  away from  $I$ .

**Definition 2.1.** A sequence of  $L^2$  - normalized bumps  $(\Phi_I)_I$  adapted to dyadic intervals  $I$  is said to be of  $\phi$  type, if and only if for each  $I$  there exists an interval  $\omega_I (= \omega_{|I|})$  symmetric with respect to the origin, so that  $\text{supp}\widehat{\Phi_I} \subseteq \omega_I$  and with  $|\omega_I| \sim |I|^{-1}$ .

**Definition 2.2.** A sequence of  $L^2$  - normalized bumps  $(\Phi_I)_I$  adapted to dyadic intervals  $I$  is said to be of  $\psi$  type, if and only if for each  $I$  there exists an interval  $\omega_I (= \omega_{|I|})$  so that  $\text{supp}\widehat{\Phi_I} \subseteq \omega_I$  and with  $|\omega_I| \sim |I|^{-1} \sim \text{dist}(0, \omega_I)$ .

Fix now  $\mathbf{n}_1, \mathbf{n}_2$  two integers and  $\mathcal{I} \subseteq \mathcal{D}$  a finite arbitrary collection of dyadic intervals. Consider also three sequences of  $L^2$  - normalized bumps  $(\Phi_{I_{\mathbf{n}_1}}^1)_{I \in \mathcal{I}}, (\Phi_{I_{\mathbf{n}_2}}^2)_{I \in \mathcal{I}}, (\Phi_I^3)_{I \in \mathcal{I}}$  adapted to  $I_{\mathbf{n}_1}, I_{\mathbf{n}_2}$  and  $I$  respectively, such that at least two of them are of  $\psi$  type.

The following theorem holds

**Theorem 2.3.** The bilinear operator defined by

$$T_{\mathcal{I}}(f, g) := \sum_{I \in \mathcal{I}} \frac{1}{|I|^{1/2}} \langle f, \Phi_{I_{\mathbf{n}_1}}^1 \rangle \langle g, \Phi_{I_{\mathbf{n}_2}}^2 \rangle \Phi_I^3 \quad (12)$$

is bounded from  $L^p \times L^q \rightarrow L^r$  for any  $1 < p, q < \infty$  and  $0 < r < \infty$  so that  $1/p + 1/q = 1/r$ , with a bound of type

$$O(\log \langle \mathbf{n}_1 \rangle \log \langle \mathbf{n}_2 \rangle)$$

depending also implicitly on  $p, q$  but independent on the cardinality of  $\mathcal{I}$  and on the families of bumps considered (here  $\langle \mathbf{n} \rangle$  simply means  $2 + |\mathbf{n}|$ ).

As we will see, this Theorem 2.3 lies at the heart of our estimates and in the rest of the section we will show how it implies the desired inequality (11). The idea will be to discretize

$C_1^{*2}$  and show that it can be reduced to operators of type (12). Equivalently, since they both have the same trilinear form, we will discretize  $C_1$  instead. We start with two Littlewood-Paley decompositions and write

$$1(\xi) = \sum_{k_1} \widehat{\Psi}_{k_1}(\xi)$$

and

$$1(\xi_1) = \sum_{k_2} \widehat{\Psi}_{k_2}(\xi_1)$$

where as usual,  $\widehat{\Psi}_{k_1}(\xi)$  and  $\widehat{\Psi}_{k_2}(\xi_1)$  are supported in the regions  $|\xi| \sim 2^{k_1}$  and  $|\xi_1| \sim 2^{k_2}$  respectively. In particular, we get

$$1(\xi, \xi_1) = \sum_{k_1, k_2} \widehat{\Psi}_{k_1}(\xi) \widehat{\Psi}_{k_2}(\xi_1). \quad (13)$$

By splitting (13) over the regions where  $k_1 \ll k_2$ ,  $k_2 \ll k_1$  and  $k_1 \sim k_2$  we obtain the final decomposition

$$1(\xi, \xi_1) = \sum_k \widehat{\Phi}_k(\xi) \widehat{\Psi}_k(\xi_1) + \quad (14)$$

$$\sum_k \widehat{\Psi}_k(\xi) \widehat{\Phi}_k(\xi_1) + \quad (15)$$

$$\sum_{k_1 \sim k_2} \widehat{\Psi}_{k_1}(\xi) \widehat{\Psi}_{k_2}(\xi_1). \quad (16)$$

By inserting this into (5),  $C_1(f, a)$  splits as a sum of three different expressions. It is easy to see that the symbol of the one corresponding to (15) is a classical symbol and for this part the inequality (11) follows from the Coifman-Meyer theorem on paraproducts [8]. We are thus left with understanding the other two terms. Notice that the first one (corresponding to (14)) interacts with the line  $\xi = 0$ , while the third one (corresponding to (16)) interacts with the line  $\xi + \xi_1 = 0$  along which the original symbol

$$\int_0^1 \text{sgn}(\xi + \alpha \xi_1) d\alpha$$

is only continuous. Also, for simplicity, from now on we will replace  $\int_0^1 \text{sgn}(\xi + \alpha \xi_1) d\alpha$  with  $\int_0^1 1_{\mathbf{R}_+}(\xi + \alpha \xi_1) d\alpha$  since the difference of the corresponding operators is just the product of  $a$  and  $f$  which clearly satisfies the original Hölder type inequalities.

Let us now fix a parameter  $k \in \mathbb{Z}$  and consider the corresponding expressions (also, since  $k_1 \sim k_2$  we assume that they are equal, for simplicity). Their trilinear forms are given by

$$\int_{\xi + \xi_1 + \xi_2 = 0} \left[ \int_0^1 1_{\mathbf{R}_+}(\xi + \alpha \xi_1) d\alpha \right] \widehat{\Phi}_k(\xi) \widehat{\Psi}_k(\xi_1) \widehat{\Psi}_k(\xi_2) \widehat{f}(\xi) \widehat{g}(\xi_1) \widehat{h}(\xi_2) d\xi d\xi_1 d\xi_2$$

and

$$\int_{\xi+\xi_1+\xi_2=0} \left[ \int_0^1 1_{\mathbf{R}_+}(\xi + \alpha\xi_1) d\alpha \right] \widehat{\Psi}_k(\xi) \widehat{\Psi}_k(\xi_1) \widehat{\Phi}_k(\xi_2) \widehat{f}(\xi) \widehat{g}(\xi_1) \widehat{h}(\xi_2) d\xi d\xi_1 d\xi_2$$

respectively. Clearly, the functions  $\widehat{\Psi}_k(\xi_2)$  and  $\widehat{\Phi}_k(\xi_2)$  have been inserted naturally into the above expressions (the first are supported away from zero while the support of the second contains the origin).

Now, on the support of  $\widehat{\Phi}_k(\xi) \widehat{\Psi}_k(\xi_1)$ , the function  $\int_0^1 1_{\mathbf{R}_+}(\xi + \alpha\xi_1) d\alpha$  can be written as a double Fourier series of type

$$\sum_{n,n_1} C_{n,n_1}^k e^{2\pi i \frac{n}{2^k} \xi} e^{2\pi i \frac{n_1}{2^k} \xi_1}. \quad (17)$$

Similarly, on the support of  $\widehat{\Psi}_k(\xi) \widehat{\Psi}_k(\xi_1)$  the same function can also be written as

$$\sum_{n,n_1} \widetilde{C}_{n,n_1}^k e^{2\pi i \frac{n}{2^k} \xi} e^{2\pi i \frac{n_1}{2^k} \xi_1}. \quad (18)$$

The following Lemma will be crucial and gives upper bounds for these Fourier coefficients.

**Lemma 2.4.** *One has*

$$|C_{n,n_1}^k| \lesssim \frac{1}{\langle n \rangle^2} \cdot \frac{1}{\langle n_1 \rangle^\#}$$

and also

$$|\widetilde{C}_{n,n_1}^k| \lesssim \frac{1}{\langle n \rangle^2} \cdot \frac{1}{\langle n - n_1 \rangle^\#} + \frac{1}{\langle n \rangle^\#} \cdot \frac{1}{\langle n_1 \rangle^\#}$$

for a fixed large integer  $\#$ , uniformly in  $k$ .

We will prove this Lemma 2.4 at the end of this section. Roughly speaking, it shows that one has at least quadratic decay for all the Fourier coefficients.

Now, (17) produces expressions of type

$$\int_{\xi+\xi_1+\xi_2=0} \left[ \widehat{\Phi}_k(\xi) e^{2\pi i \frac{n}{2^k} \xi} \right] \left[ \widehat{\Psi}_k(\xi_1) e^{2\pi i \frac{n_1}{2^k} \xi_1} \right] \widehat{\Psi}_k(\xi_2) \widehat{f}(\xi) \widehat{g}(\xi_1) \widehat{h}(\xi_2) d\xi d\xi_1 d\xi_2$$

which can be rewritten as

$$\int_{\mathbf{R}} (f * \Phi_k^{1,n})(x) (g * \Psi_k^{2,n_1})(x) (h * \Psi_k^3)(x) dx$$

and this can be further discretized by standard arguments as an average of expressions of type

$$\sum_{|I|=2^{-k}} \frac{1}{|I|^{1/2}} \langle f, \Phi_{I_n}^1 \rangle \langle g, \Phi_{I_{n_1}}^2 \rangle \Phi_I^3, \quad (19)$$

where the functions  $\Phi_{I_{n_1}}^2$  and  $\Phi_I^3$  are of  $\psi$  type.

Similarly, (18) produces expressions of type

$$\int_{\xi+\xi_1+\xi_2=0} \left[ \widehat{\Psi}_k(\xi) e^{2\pi i \frac{n}{2^k} \xi} \right] \left[ \widehat{\Psi}_k(\xi_1) e^{2\pi i \frac{n_1}{2^k} \xi_1} \right] \widehat{\Phi}_k(\xi_2) \widehat{f}(\xi) \widehat{g}(\xi_1) \widehat{h}(\xi_2) d\xi d\xi_1 d\xi_2$$

and as we have seen, these can be further rewritten and discretized again in the form (19), where this time  $\Phi_{I_n}^1$  and  $\Phi_{I_{n_1}}^2$  are of  $\psi$  type. The connection with (12) should be clear by now. If we add all the expressions of type (19) for all the scales  $k \in \mathbb{Z}$ , one obtains a discrete trilinear form corresponding to the part of  $C_1$  related to (14) (and of course, as we discussed, there is a similar one related to (16)). In particular, since we are interested in estimating  $C_1^{*2}$ , its bilinear model is of the form

$$\sum_{I \in \mathcal{I}} \frac{1}{|I|^{1/2}} \langle f, \Phi_{I_n}^1 \rangle \langle h, \Phi_{I_{n_1}}^2 \rangle \Phi_{I_{n_1}}^3$$

which should be rewritten as

$$\sum_{I \in \mathcal{I}} \frac{1}{|I|^{1/2}} \langle f, \Phi_{I_{n-n_1}}^1 \rangle \langle h, \Phi_{I_{-n_1}}^2 \rangle \Phi_I^3$$

to be able to compare it better with (12).

Now, using the fact that  $C_1^{*2}(f, g)$  makes perfect sense for Schwartz functions (in fact, it can be written as an expression similar to (5)) and by triangle inequality, Fatou's lemma and Theorem 2.3, it follows that

$$\|C_1^{*2}(f, g)\|_1 \lesssim \sum_{n, n_1} \sup_k (|C_{n, n_1}^k|, |\widetilde{C_{n, n_1}^k}|) \cdot \log < n - n_1 > \cdot \log < n_1 > \cdot \|f\|_p \cdot \|g\|_{p'}$$

which is clearly bounded by  $\|f\|_p \cdot \|g\|_{p'}$  as a consequence of the quadratic decay in Lemma 2.4. This completes the proof of (11).

We now describe the proof of Lemma 2.4. We first record the following

**Lemma 2.5.** *One has the following identities*

- (a)  $\partial_\xi^2 \left( \int_0^{\xi_1} 1_{\mathbf{R}_+}(\xi + \alpha) d\alpha \right) = \delta_0(\xi + \xi_1) - \delta_0(\xi).$
  - (b)  $\partial_{\xi_1}^2 \left( \int_0^{\xi_1} 1_{\mathbf{R}_+}(\xi + \alpha) d\alpha \right) = \delta_0(\xi + \xi_1).$
  - (c)  $\partial_\xi \partial_{\xi_1} \left( \int_0^{\xi_1} 1_{\mathbf{R}_+}(\xi + \alpha) d\alpha \right) = \partial_{\xi_1} \partial_\xi \left( \int_0^{\xi_1} 1_{\mathbf{R}_+}(\xi + \alpha) d\alpha \right) = \delta_0(\xi + \xi_1).$
- where  $\delta_0$  is the Dirac distribution with respect to the origin.

**Proof**

This is really straightforward. Let us verify (a) for example. One has

$$\partial_\xi^2 \left( \int_0^{\xi_1} 1_{\mathbf{R}_+}(\xi + \alpha) d\alpha \right) = \partial_\xi \left( \int_0^{\xi_1} \delta_0(\xi + \alpha) d\alpha \right) = \partial_\xi \left( \int_\xi^{\xi+\xi_1} \delta_0(\alpha) d\alpha \right) = \delta_0(\xi + \xi_1) - \delta_0(\xi).$$

■

To prove now the estimates in Lemma 2.4, we rewrite (for instance)  $\widetilde{C_{n, n_1}^k}$  as

$$\frac{1}{2^k} \frac{1}{2^k} \int_{\mathbf{R}^2} \left[ \int_0^1 1_{\mathbf{R}_+}(\xi + \alpha \xi_1) d\alpha \right] \widehat{\Psi}_k(\xi) \widehat{\Psi}_k(\xi_1) e^{-2\pi i \frac{n}{2^k} \xi} e^{-2\pi i \frac{n_1}{2^k} \xi_1} d\xi d\xi_1 =$$



$$\begin{aligned}
& \int_{\mathbf{R}^2} \left[ \int_0^1 1_{\mathbf{R}_+}(\xi + \alpha \xi_1) d\alpha \right] \widehat{\Psi}(\xi) \widehat{\Psi}(\xi_1) e^{-2\pi i n \xi} e^{-2\pi i n_1 \xi_1} d\xi d\xi_1 = \\
& \int_{\mathbf{R}^2} \left[ \frac{1}{\xi_1} \int_0^{\xi_1} 1_{\mathbf{R}_+}(\xi + \alpha) d\alpha \right] \widehat{\Psi}(\xi) \widehat{\Psi}(\xi_1) e^{-2\pi i n \xi} e^{-2\pi i n_1 \xi_1} d\xi d\xi_1 := \\
& \int_{\mathbf{R}^2} \left[ \int_0^{\xi_1} 1_{\mathbf{R}_+}(\xi + \alpha) d\alpha \right] \widehat{\Psi}(\xi) \widehat{\widetilde{\Psi}}(\xi_1) e^{-2\pi i n \xi} e^{-2\pi i n_1 \xi_1} d\xi d\xi_1, \tag{20}
\end{aligned}$$

where  $\widehat{\Psi}(\xi)$ ,  $\widehat{\Psi}(\xi_1)$  and  $\widehat{\widetilde{\Psi}}(\xi_1)$  are supported away from the origin and are adapted to scale 1.

The idea is of course to integrate by parts as much as we can in (20) and keep track of the upper bounds that one gets in this way. We begin integrating by parts in  $\xi$  as much as we can. Since both  $\int_0^{\xi_1} 1_{\mathbf{R}_+}(\xi + \alpha) d\alpha$  and  $\widehat{\Psi}(\xi)$  depend on  $\xi$ , the  $\xi$  derivatives can hit either of the terms. If the derivative hits twice the term  $\int_0^{\xi_1} 1_{\mathbf{R}_+}(\xi + \alpha) d\alpha$  then, because of the previous Lemma 2.5, the  $\xi$  variable disappears and becomes  $-\xi_1$  (notice that  $\xi$  cannot be zero in this case) at which point (20) gets simplified into an expression of type

$$\int_{\mathbf{R}} \widehat{\Psi}(-\xi_1) \widehat{\widetilde{\Psi}}(\xi_1) e^{-2\pi i \xi_1 (n-n_1)} d\xi_1.$$

But this term can be integrated by parts as many times as we wish and this explains the appearance of the first upper bound for  $|\widetilde{C}_{n,n_1}^k|$ . If on the contrary, the  $\xi$  derivative didn't hit the term  $\int_0^{\xi_1} 1_{\mathbf{R}_+}(\xi + \alpha) d\alpha$  two times yet even after many integration by parts, this means that we already gained a factor of type  $\frac{1}{\langle n \rangle^\#}$  at which moment we stop integrating in  $\xi$  and start integrating by parts in  $\xi_1$ . As before, if the  $\xi_1$  derivatives hit the term  $\int_0^{\xi_1} 1_{\mathbf{R}_+}(\xi + \alpha) d\alpha$  until one reaches  $\delta_0(\xi + \xi_1)$  then  $\xi_1$  becomes  $-\xi$  and after that one integrates by parts a smooth function obtaining an upper bound of type  $\frac{1}{\langle n \rangle^\#} \frac{1}{\langle n-n_1 \rangle^\#}$  which is smaller than the previously discussed one.

If finally, the  $\xi_1$  derivative didn't hit  $\int_0^{\xi_1} 1_{\mathbf{R}_+}(\xi + \alpha) d\alpha$  until it becomes  $\delta_0(\xi + \xi_1)$ , then this means that it keeps hitting the smooth function of  $\xi_1$  in which case we obtain an upper bound of type  $\frac{1}{\langle n \rangle^\#} \frac{1}{\langle n_1 \rangle^\#}$  as desired.

The second term  $C_{n,n_1}^k$  can be treated similarly. One should just remark that in this case the equality  $\xi_1 = -\xi$  is impossible and only  $\delta_0(\xi)$  remains after integrating by parts, which explains the slight difference between the two upper bounds.

We are as a consequence left with proving Theorem 2.3.

### 3. PROOF OF THEOREM 2.3

The proof is based on the method introduced in [21] and [22].

Let us assume without loss of generality that the families  $(\Phi_{I_{n_2}}^2)_I$  and  $(\Phi_I^3)_I$  are of  $\psi$  type (since all the other possible cases can be treated in a similar way). Fix also  $1 < p, q < \infty$  and  $0 < r < \infty$  so that  $1/p + 1/q = 1/r$ . We will prove that  $T_I$  maps  $L^p \times L^q \rightarrow L^{r,\infty}$  since then (12) follows easily by standard interpolation arguments.

As usual, it is enough to show that given a measurable set  $E \subseteq \mathbf{R}$  with  $|E| = 1$ , one can find  $E' \subseteq E$  with  $|E'| \sim 1$  and so that

$$\sum_{I \in I} \frac{1}{|I|^{1/2}} |\langle f, \Phi_{I_{n_1}}^1 \rangle| |\langle g, \Phi_{I_{n_2}}^2 \rangle| |\langle h, \Phi_I^3 \rangle| \lesssim (\log \langle \mathbf{n}_1 \rangle) (\log \langle \mathbf{n}_2 \rangle) \quad (21)$$

where  $h := \chi_{E'}$ .

Define now the shifted maximal operator  $M^{\mathbf{n}_1}$  and the shifted square function  $S^{\mathbf{n}_2}$  as follows.

$$M^{\mathbf{n}_1} f(x) := \sup_{x \in I} \frac{1}{|I|} \int_{\mathbf{R}} |f(y)| \widetilde{\chi_{I_{n_1}}}(y) dy$$

where  $\widetilde{\chi_{I_{n_1}}}(y)$  denotes the function

$$\widetilde{\chi_{I_{n_1}}}(y) = \left( 1 + \frac{\text{dist}(y, I_{n_1})}{|I_{n_1}|} \right)^{-100},$$

while  $S^{\mathbf{n}_2}$  is given by

$$S^{\mathbf{n}_2} g(x) := \left( \sum_I \frac{|\langle g, \Phi_{I_{n_2}}^2 \rangle|^2}{|I|} 1_I(x) \right)^{1/2}.$$

As we will see later on, both of them are bounded on every  $L^p$  space for  $1 < p < \infty$ , with a bound of type  $O(\log \langle \mathbf{n}_1 \rangle)$  and  $O(\log \langle \mathbf{n}_2 \rangle)$  respectively. Using these two facts we define an exceptional set as follows.

First, define the set  $\widetilde{\Omega}_0$  by

$$\widetilde{\Omega}_0 := \{x : M^{\mathbf{n}_1} f(x) > C \log \langle \mathbf{n}_1 \rangle\} \cup \{x : S^{\mathbf{n}_2} f(x) > C \log \langle \mathbf{n}_2 \rangle\}.$$

Let now  $d$  a positive integer and  $\#$  an integer so that  $2^d < \# \leq 2^{d+1}$ . Define the set  $\Omega_{\#}^d$  by

$$\Omega_{\#}^d := \left\{ x : M^{\mathbf{n}_1 - \#} f(x) > C \log \langle \mathbf{n}_1 - \# \rangle 2^{5d} \right\}$$

and then the set  $\widetilde{\widetilde{\Omega}}_0$  by

$$\widetilde{\widetilde{\Omega}}_0 := \bigcup_{d \geq 0} \bigcup_{2^d < \# \leq 2^{d+1}} \Omega_{\#}^d.$$

Define also the set  $\widetilde{\widetilde{\widetilde{\Omega}}}_0$  in a similar way to  $\widetilde{\widetilde{\Omega}}_0$  but by using the function  $g$  and the corresponding index  $\mathbf{n}_2$  instead. Then, define  $\Omega_0$  to be

$$\Omega_0 := \widetilde{\Omega}_0 \cup \widetilde{\widetilde{\Omega}}_0 \cup \widetilde{\widetilde{\widetilde{\Omega}}}_0 \quad (22)$$

and finally, the exceptional set

$$\Omega := \left\{ x : M(1_{\Omega_0})(x) > \frac{1}{100} \right\}.$$

Observe that  $|\Omega| \ll 1$  if  $C$  is chosen large enough and this allows us to define the set  $E'$  by  $E' := E \setminus \Omega$  and to observe that  $|E'| \sim 1$  as desired. To be able to estimate (21) properly, we split it into two parts as follows:

$$\sum_{I \cap \Omega^c \neq \emptyset} + \sum_{I \cap \Omega^c = \emptyset} := I + II. \quad (23)$$

**Estimates for I.** First, we observe that since  $I \cap \Omega^c \neq \emptyset$  one has  $\frac{|I \cap \Omega_0|}{|I|} < \frac{1}{100}$  which means that  $|I \cap \Omega_0^c| > \frac{99}{100}|I|$ .

We now perform three independent stopping time type arguments for the functions  $f, g, h$  which will be combined carefully later on.

Define first

$$\Omega_1 = \left\{ x : M^{\mathbf{n}_1}(f)(x) > \frac{C \log < \mathbf{n}_1 >}{2^1} \right\}$$

and set

$$\mathcal{I}_1 = \left\{ I \in \mathcal{I} : |I \cap \Omega_1| > \frac{1}{100}|I| \right\},$$

then define

$$\Omega_2 = \left\{ x : M^{\mathbf{n}_1}(f)(x) > \frac{C \log < \mathbf{n}_1 >}{2^2} \right\}$$

and set

$$\mathcal{I}_2 = \left\{ I \in \mathcal{I} \setminus \mathcal{I}_1 : |I \cap \Omega_2| > \frac{1}{100}|I| \right\},$$

and so on. The constant  $C > 0$  here is the one in the definition of the set  $E'$  before. Clearly, since  $\mathcal{I}$  is finite, we will run out of dyadic intervals after a while, thus producing the sets  $(\{\Omega_n\})_n$  and  $(\{\mathcal{I}_n\})_n$ .

Independently, define

$$\Omega'_1 = \left\{ x : S^{\mathbf{n}_2}(g)(x) > \frac{C \log < n_2 >}{2^1} \right\}$$

and set

$$\mathcal{I}'_1 = \left\{ I \in \mathcal{I} : |I \cap \Omega'_1| > \frac{1}{100}|I| \right\},$$

then as before define

$$\Omega'_2 = \left\{ x : S^{\mathbf{n}_2}(g)(x) > \frac{C \log < n_2 >}{2^2} \right\}$$

and set

$$\mathcal{I}'_2 = \left\{ I \in \mathcal{I} \setminus \mathcal{I}'_1 : |I \cap \Omega'_2| > \frac{1}{100}|I| \right\},$$

and so on, producing the finitely many sets  $(\{\Omega'_n\})_n$  and  $(\{\mathcal{I}'_n\})_n$ . Of course, we would like to have such a decomposition available for  $h$  as well. To do this, we first need to construct the

analogue of the set  $\Omega_0$ , for it. To do this, first pick  $N > 0$  a big enough integer such that for every  $I \in \mathcal{I}$  we have  $|I \cap \Omega_{-N}''^c| > \frac{99}{100}|I|$  where we defined

$$\Omega_{-N}'' = \left\{x : S(h)(x) > C2^N\right\}.$$

Then, similarly to the previous algorithms, we define

$$\Omega_{-N+1}'' = \left\{x : S(h)(x) > \frac{C2^N}{2^1}\right\}$$

and set

$$\mathcal{I}_{-N+1}'' = \left\{I \in \mathcal{I} : |I \cap \Omega_{-N+1}''| > \frac{1}{100}|I|\right\},$$

then define

$$\Omega_{-N+2}'' = \left\{x : S(h)(x) > \frac{C2^N}{2^2}\right\}$$

and set

$$\mathcal{I}_{-N+2}'' = \left\{I \in \mathcal{I} \setminus \mathcal{I}_{-N+1}'' : |I \cap \Omega_{-N+2}''| > \frac{1}{100}|I|\right\},$$

and so on, constructing the finitely many sets  $(\{\Omega_n''\})_n$  and  $(\{\mathcal{I}_n''\})_n$ .

Using all these decompositions, we can decompose term  $I$  further as

$$\sum_{l_1, l_2 > 0, l_3 > -N} \sum_{I \in \mathcal{I}_{l_1, l_2, l_3}} \frac{1}{|I|^{3/2}} |\langle f, \Phi_{I_{n_1}}^1 \rangle| |\langle g, \Phi_{I_{n_2}}^2 \rangle| |\langle h, \Phi_I^3 \rangle| |I|, \quad (24)$$

where

$$\mathcal{I}_{l_1, l_2, l_3} := \mathcal{I}_{l_1} \cap \mathcal{I}_{l_2}' \cap \mathcal{I}_{l_3}''.$$

Then, observe that since  $I$  belongs to  $\mathcal{I}_{l_1, l_2, l_3}$  this means in particular that  $I$  has not been selected at either of the previous  $l_1 - 1$ ,  $l_2 - 1$  and  $l_3 - 1$  steps respectively, which means that all of  $|I \cap \Omega_{l_1-1}|$ ,  $|I \cap \Omega_{l_2-1}'|$  and  $|I \cap \Omega_{l_3-1}''|$  are smaller than  $\frac{1}{100}|I|$  or equivalently, that one has

$$|I \cap \Omega_{l_1-1}^c| > \frac{99}{100}|I|,$$

$$|I \cap \Omega_{l_2-1}'^c| > \frac{99}{100}|I|$$

and

$$|I \cap \Omega_{l_3-1}''^c| > \frac{99}{100}|I|,$$

which implies that

$$|I \cap \Omega_{l_1-1}^c \cap \Omega_{l_2-1}'^c \cap \Omega_{l_3-1}''^c| > \frac{97}{100}|I|. \quad (25)$$

Using this in (24) one can estimate that expression by

$$\begin{aligned}
& \sum_{l_1, l_2 > 0, l_3 > -N} \sum_{I \in \mathcal{I}_{l_1, l_2, l_3}} \frac{1}{|I|^{3/2}} |\langle f, \Phi_{I_{n_1}}^1 \rangle| |\langle g, \Phi_{I_{n_2}}^2 \rangle| |\langle h, \Phi_I^3 \rangle| |I \cap \Omega_{l_1-1}^c \cap \Omega_{l_2-1}^{c'} \cap \Omega_{l_3-1}^{c''}| = \\
& \sum_{l_1, l_2 > 0, l_3 > -N} \int_{\Omega_{l_1-1}^c \cap \Omega_{l_2-1}^{c'} \cap \Omega_{l_3-1}^{c''}} \sum_{I \in \mathcal{I}_{l_1, l_2, l_3}} \frac{|\langle f, \Phi_{I_{n_1}}^1 \rangle| |\langle g, \Phi_{I_{n_2}}^2 \rangle| |\langle h, \Phi_I^3 \rangle|}{|I|^{1/2} |I|^{1/2} |I|^{1/2}} \chi_I(x) dx \\
& \lesssim \sum_{l_1, l_2 > 0, l_3 > -N} \int_{\Omega_{l_1-1}^c \cap \Omega_{l_2-1}^{c'} \cap \Omega_{l_3-1}^{c''} \cap \Omega_{I_{l_1, l_2, l_3}}} M^{n_1}(f)(x) S^{n_2}(g)(x) S(h)(x) dx \\
& \lesssim \sum_{l_1, l_2 > 0, l_3 > -N} \log < \mathbf{n}_1 > \log < \mathbf{n}_2 > 2^{-l_1} 2^{-l_2} 2^{-l_3} |\Omega_{I_{l_1, l_2, l_3}}|, \tag{26}
\end{aligned}$$

where

$$\Omega_{I_{l_1, l_2, l_3}} := \bigcup_{I \in \mathcal{I}_{l_1, l_2, l_3}} I.$$

On the other hand, we also have

$$\begin{aligned}
|\Omega_{I_{l_1, l_2, l_3}}| & \leq |\Omega_{I_{l_1}}| \leq \left| \left\{ x : M(\chi_{\Omega_{l_1}})(x) > \frac{1}{100} \right\} \right| \\
& \lesssim |\Omega_{l_1}| = \left| \left\{ x : M^{n_1}(f)(x) > \frac{C \log < n_1 >}{2^{l_1}} \right\} \right| \lesssim 2^{l_1 p}.
\end{aligned}$$

Similarly, we have

$$|\Omega_{I_{l_1, l_2, l_3}}| \lesssim 2^{l_2 q}$$

and also

$$|\Omega_{I_{l_1, l_2, l_3}}| \lesssim 2^{l_3 \alpha},$$

for every  $\alpha > 1$ . Here we used the fact that all the operators  $M^{n_1}$ ,  $S^{n_2}$  and  $S$  are bounded on  $L^s$  as long as  $1 < s < \infty$  and also that  $|E'_3| \sim 1$ . In particular, this implies that

$$|\Omega_{I_{l_1, l_2, l_3}}| \lesssim 2^{l_1 p \theta_1} 2^{l_2 q \theta_2} 2^{l_3 \alpha \theta_3} \tag{27}$$

for any  $0 \leq \theta_1, \theta_2, \theta_3 < 1$ , such that  $\theta_1 + \theta_2 + \theta_3 = 1$ .

On the other hand, (26) can be split into

$$\log < \mathbf{n}_1 > \log < \mathbf{n}_2 > \left( \sum_{l_1, l_2 > 0, l_3 > 0} 2^{-l_1} 2^{-l_2} 2^{-l_3} |\Omega_{I_{l_1, l_2, l_3}}| + \sum_{l_1, l_2 > 0, 0 > l_3 > -N} 2^{-l_1} 2^{-l_2} 2^{-l_3} |\Omega_{I_{l_1, l_2, l_3}}| \right). \tag{28}$$

To estimate the first expression in (28) we use the inequality (27) for  $\theta_1, \theta_2, \theta_3$  so that  $1 - p\theta_1 > 0$ ,  $1 - q\theta_2 > 0$  and  $1 - \alpha\theta_3 > 0$ , while to estimate the second term we use (27) for  $\theta_1, \theta_2, \theta_3$  such that  $1 - p\theta_1 > 0$ ,  $1 - q\theta_2 > 0$  and  $1 - \alpha\theta_3 < 0$ . With these choices, the sum in (28) is indeed is  $O(\log < \mathbf{n}_1 > \log < \mathbf{n}_2 >)$  as desired. This ends the discussion of  $I$ .

**Estimates for II.** This term is simpler to estimate, now that we defined our exceptional set so carefully. Notice that the intervals of interest are those inside  $\Omega$ . One can split them as  $\bigcup_{d \geq 0} \mathcal{I}_d$  where

$$\mathcal{I}_d := \left\{ I \in \mathcal{I} : I \subseteq \Omega \text{ and } 2^d \leq \frac{\text{dist}(I, \Omega^c)}{|I|} < 2^{d+1} \right\}.$$

Observe that for any  $d \geq 0$  one has

$$\sum_{I \in \mathcal{I}_d} |I| \lesssim |\Omega| \lesssim 1.$$

Also, for every  $I \in \mathcal{I}_d$  one has that  $2^d I \cap \Omega^c = \emptyset$  and also there exists  $\widetilde{I}$  dyadic and of the same length, which lies  $\#$  steps of length  $|I|$  away from  $I$  (with  $2^d \leq \# \leq 2^{d+1}$ ), and having the property that  $\widetilde{I} \cap \Omega^c \neq \emptyset$ . In particular, this means that  $I_{n_1}$  and  $I_{n_2}$  are  $n_1 - \#$  and  $n_2 - \#$  steps of length  $|I|$  away from  $\widetilde{I}$ . Using all these facts, one can estimate term  $II$  by

$$\begin{aligned} & \sum_{d \geq 0} \sum_{I \in \mathcal{I}_d} \frac{|\langle f, \Phi_{I_{n_1}}^1 \rangle| |\langle g, \Phi_{I_{n_2}}^2 \rangle| |\langle h, \Phi_I^3 \rangle|}{|I|^{1/2} |I|^{1/2} |I|^{1/2}} |I| \lesssim \\ & \sum_{d \geq 0} \sum_{2^d \leq \# \leq 2^{d+1}} \sum_{I \in \mathcal{I}_d} (\log \langle n_1 - \# \rangle) 2^{5d} (\log \langle n_2 - \# \rangle) 2^{5d} 2^{-Md} |I| \lesssim \\ & (\log \langle n_1 \rangle) (\log \langle n_2 \rangle) \end{aligned} \tag{29}$$

by using the trivial fact that

$$\log \langle n_j - \# \rangle \leq \log \langle n_j \rangle + \langle \# \rangle$$

for  $j = 1, 2$ .

The proof is now complete.

#### 4. APPENDIX 1: LOGARITHMIC ESTIMATES FOR THE SHIFTED MAXIMAL FUNCTION

The goal of this section is to prove the following theorem that have been used before. This result can be found in Stein [25], but we decided to give a selfcontained proof of it here (which we (re)discovered independently), not only for reader's convenience, but also for the fact that some particular notations that will be introduced, will be quoted and used later on as well.

**Theorem 4.1.** ([25]) *For any  $\mathbf{n} \in \mathbb{Z}$ , the shifted maximal function  $M^{\mathbf{n}}$  is bounded on every  $L^p$  space for  $1 < p < \infty$ , with a bound of type  $O(\log \langle \mathbf{n} \rangle)$ .*

**Proof** First, we observe that in order to prove the desired estimates, it is enough to prove them for the corresponding *sharp* maximal function  $\widetilde{M}^{\mathbf{n}}$  defined by

$$\widetilde{M}^{\mathbf{n}} f(x) := \sup_{x \in I} \frac{1}{|I_{\mathbf{n}}|} \int_{I_{\mathbf{n}}} |f(y)| dy \tag{30}$$

where the supremum is taken only over dyadic intervals.

To see this, fix  $x$  and  $I$  so that  $x \in I$ . One can write

$$\frac{1}{|I_n|} \int_{I_n} |f(y)| dy \lesssim \sum_{\# \in \mathbb{Z}} \left[ \frac{1}{|I_n^\#|} \int_{I_n^\#} |f(y)| dy \right] \frac{1}{\langle \# \rangle^{100}},$$

where  $I_n^\#$  is the dyadic interval of the same length with  $I_n$  and lying  $\#$  steps of length  $|I_n|$  away from it. In particular, using the above and assuming that the theorem holds for  $\widetilde{M}^n$ , one has

$$\begin{aligned} \|M^n f\|_p &\lesssim \sum_{\# \in \mathbb{Z}} \frac{1}{\langle \# \rangle^{100}} \|\widetilde{M}^{n+\#} f\|_p \lesssim \sum_{\# \in \mathbb{Z}} \frac{1}{\langle \# \rangle^{100}} (\log \langle n + \# \rangle) \|f\|_p \lesssim \\ &\lesssim \sum_{\# \in \mathbb{Z}} \frac{1}{\langle \# \rangle^{100}} (\log(\langle n \rangle \langle \# \rangle)) \lesssim \log \langle n \rangle \|f\|_p, \end{aligned}$$

as desired. We are then left with proving the theorem for  $\widetilde{M}^n$ .

Let now  $\lambda > 0$ . We claim that one has the following inequality

$$|\{x : \widetilde{M}^n f(x) > \lambda\}| \lesssim (\log \langle n \rangle) |\{x : Mf(x) > \lambda\}| \quad (31)$$

where  $M$  is the classical Hardy-Littlewood maximal operator. Assuming (31), the theorem for  $\widetilde{M}^n$  follows from the Hardy-Littlewood theorem by interpolation with the trivial  $L^\infty$  estimate.

To finally prove (31) denote by  $I_n^\lambda$  the collection of all dyadic and maximal with respect to inclusion intervals  $I_n$ , for which

$$\frac{1}{|I_n|} \int_{I_n} |f(y)| dy > \lambda.$$

Note that all of them are disjoint and one also has

$$\bigcup_{I_n \in I_n^\lambda} I_n = \{x : Mf(x) > \lambda\}.$$

Then, for every such a selected maximal dyadic interval  $I_n$ , consider its dyadic subintervals of length  $|I_n|$ ,  $|I_n|/2$ ,  $|I_n|/2^2$  ..., etc. Observe that there exist only  $\lceil \log \langle n \rangle \rceil$  disjoint dyadic intervals  $I_n^1, I_n^2, \dots, I_n^{\lceil \log \langle n \rangle \rceil}$  of the same length with  $|I_n|$ , so that the translate with  $-n$  corresponding units of any such smaller dyadic subinterval of  $I_n$ , becomes a subinterval of one of these  $I_n^1, I_n^2, \dots, I_n^{\lceil \log \langle n \rangle \rceil}$ . The claim is now that

$$\{x : \widetilde{M}^n f(x) > \lambda\} \subseteq \bigcup_{I_n \in I_n^\lambda} (I_n \cup I_n^1 \cup \dots \cup I_n^{\lceil \log \langle n \rangle \rceil}).$$

To see this, pick  $x_0$  so that  $M^n f(x_0) > \lambda$ . Then, this means that there exists a dyadic interval  $J$  containing  $x_0$ , so that  $\frac{1}{|J_n|} \int_{J_n} |f(y)| dy > \lambda$ . Because of the previous construction, one can for sure find one selected maximal interval of type  $I_n$ , so that  $J_n \subseteq I_n$ . But then, this means in particular that  $J$  itself will be a subset of either  $I_n$  or  $I_n^1$  or ... or  $I_n^{\lceil \log \langle n \rangle \rceil}$  which implies the claim.

It is now easy to see that this claim together with the disjointness of the maximal intervals  $I_n$ , imply (31). The proof is then complete. <sup>3</sup>

■

<sup>3</sup>Of course, since the trivial  $L^\infty$  estimate comes with an  $O(1)$  bound, by interpolation the  $L^p$  operatorial bound of  $M^n$  will be even  $O((\log \langle n \rangle)^{1/p})$ . But for simplicity, we used the  $O(\log \langle n \rangle)$  bound all the time.

## 5. APPENDIX 2: LOGARITHMIC ESTIMATES FOR THE SHIFTED SQUARE FUNCTION

The goal of this last section is to prove the following theorem which played an important role earlier in the argument <sup>4</sup>.

**Theorem 5.1.** *For any  $n \in \mathbb{Z}$ , the shifted square function  $S^n$  is bounded on every  $L^p$  space for  $1 < p < \infty$ , with a bound of type  $O(\log < n >)$ .*

**Proof**

Besides the observations of the previous section, the proof is based on a classical Calderón-Zygmund decomposition [25].

First, let us observe that  $S^n$  is bounded on  $L^2$  with a bound independent of  $n$ . Indeed, one can see that

$$\|S^n f\|_2 = \left( \sum_I \langle f, \Phi_{I_n} \rangle^2 \right)^{1/2}$$

which is clearly comparable to the  $L^2$  norm of the classical Littlewood-Paley square function, which is known to be bounded on  $L^2$ .

Next, we show that

$$\|S^n f\|_{1,\infty} \lesssim (\log < n >) \|f\|_1. \quad (32)$$

or more specifically that

$$|\{x \in \mathbf{R} : S^n f(x) > \lambda\}| \lesssim \log < n > \frac{1}{\lambda} \|f\|_1. \quad (33)$$

Fix such a  $\lambda > 0$  and perform a Calderón-Zygmund decomposition of the function  $f$  at level  $\lambda$ . Pick one by one maximal dyadic intervals  $J$  so that

$$\frac{1}{|J|} \int_J |f(y)| dy > \lambda.$$

Observe that all these intervals are by construction disjoint and denote their union with  $\Omega$ . One has

$$|\Omega| = \sum_J |J| < \frac{1}{\lambda} \sum_J \int_J |f(y)| dy \leq \frac{1}{\lambda} \|f\|_1. \quad (34)$$

Split now the function  $f$  as

$$f = g + b$$

where

$$g := f \chi_{\Omega^c} + \sum_J \left[ \frac{1}{|J|} \int_J f(y) dy \right] \chi_J$$

---

<sup>4</sup>It may very well be that this result has been observed before (as it was the case with the previous shifted maximal function) but since we didn't find it in the literature, we include a selfcontained proof of it in what follows.



and

$$b := f - g := \sum_J b_J$$

where

$$b_J := \left[ f - \frac{1}{|J|} \int_J f(y) dy \right] \chi_J.$$

Clearly,  $\text{supp} b_J \subseteq J$ . Observe also that one has

$$|f(x)| \leq \lambda$$

for every  $x \in \Omega^c$  and as a consequence,

$$\|g\|_\infty \lesssim \lambda$$

since one also observes that

$$\left| \frac{1}{|J|} \int_J f(y) dy \right| \leq \frac{1}{|J|} \int_J |f(y)| dy \leq \frac{2}{|\tilde{J}|} \int_{\tilde{J}} |f(y)| dy \leq 2\lambda$$

where  $\tilde{J}$  is the unique dyadic interval containing  $J$  and twice as long. It is also important to observe that

$$\int_{\mathbf{R}} b_J(y) dy = 0$$

by definition and also that

$$\|b_J\|_1 = \int_J |b_J(y)| dy \leq \int_J |f(y)| dy + \left( \frac{1}{|J|} \int_J |f(y)| dy \right) |J| \lesssim \int_J |f(y)| dy \lesssim \lambda |J|$$

as we have seen.

Using all these properties, one can write

$$|\{x \in \mathbf{R} : S^n f(x) > \lambda\}| \leq |\{x \in \mathbf{R} : S^n g(x) > \lambda/2\}| + |\{x \in \mathbf{R} : S^n b(x) > \lambda/2\}|. \quad (35)$$

To estimate the first term in (35), we use the  $L^2$  boundedness of  $S^n$  and we write

$$\begin{aligned} |\{x \in \mathbf{R} : S^n g(x) > \lambda/2\}| &\lesssim \frac{1}{\lambda^2} \|S^n g\|_2^2 \lesssim \\ \frac{1}{\lambda^2} \|g\|_2^2 &= \frac{1}{\lambda^2} \int_{\mathbf{R}} |g(x)|^2 dx \lesssim \frac{1}{\lambda^2} \lambda \int_{\mathbf{R}} |g(x)| dx = \\ \frac{1}{\lambda} \|g\|_1 &\lesssim \frac{1}{\lambda} \left( \int_{\Omega^c} |f(x)| dx + \sum_J \int_J |f(x)| dx \right) \lesssim \frac{1}{\lambda} \|f\|_1, \end{aligned}$$

as desired.

To estimate the second term in (35), we proceed as follows. First, for any interval  $J$ , consider the associated  $J^1, J^2, \dots, J^{[\log \langle n \rangle]}$  as defined in the previous section and define the set  $\Omega_J$  by

$$\Omega_J := 5J \cup 5J^1 \cup 5J^2 \cup \dots \cup 5J^{\lceil \log \langle \mathbf{n} \rangle \rceil}.$$

Then, one has

$$\begin{aligned} |\{x \in \mathbf{R} : S^n b(x) > \lambda/2\}| &\leq \left| \left\{ x \in \bigcup_J \Omega_J : S^n b(x) > \lambda/2 \right\} \right| + \\ &\quad \left| \left\{ x \in \left( \bigcup_J \Omega_J \right)^c : S^n b(x) > \lambda/2 \right\} \right|. \end{aligned} \quad (36)$$

The first expression is easy to estimate since one can write

$$\left| \left\{ x \in \bigcup_J \Omega_J : S^n b(x) > \lambda/2 \right\} \right| \leq \left| \bigcup_J \Omega_J \right| \lesssim (\log \langle \mathbf{n} \rangle) \sum_J |J| \lesssim (\log \langle \mathbf{n} \rangle) \frac{1}{\lambda} \|f\|_1,$$

as we have seen before. The second expression in (36) can be majorized by

$$\frac{1}{\lambda} \int_{(\bigcup_J \Omega_J)^c} S^n b(x) dx \leq \frac{1}{\lambda} \sum_J \int_{(\bigcup_J \Omega_J)^c} S^n b_J(x) dx \leq \frac{1}{\lambda} \sum_J \int_{(\Omega_J)^c} S^n b_J(x) dx$$

and we claim now that for any  $J$  one has

$$\int_{(\Omega_J)^c} S^n b_J(x) dx \lesssim \lambda |J|. \quad (37)$$

Assuming (37), one can continue the previous inequality and further majorize it by

$$\frac{1}{\lambda} \lambda \sum_J |J| \lesssim |\Omega| \lesssim \frac{1}{\lambda} \|f\|_1$$

as desired.

We are then left with proving our claim (37). First, we majorize the left hand side of it by

$$\begin{aligned} \int_{(\Omega_J)^c} \left( \sum_I \frac{|\langle b_J, \Phi_{I_n} \rangle|}{|I|^{1/2}} 1_I(x) \right) dx &= \sum_I \int_{(\Omega_J)^c} \frac{|\langle b_J, \Phi_{I_n} \rangle|}{|I|^{1/2}} 1_I(x) dx = \\ &= \sum_{|I| \leq |J|} \int_{(\Omega_J)^c} \frac{|\langle b_J, \Phi_{I_n} \rangle|}{|I|^{1/2}} 1_I(x) dx + \sum_{|I| > |J|} \int_{(\Omega_J)^c} \frac{|\langle b_J, \Phi_{I_n} \rangle|}{|I|^{1/2}} 1_I(x) dx := \\ &= A + B. \end{aligned}$$

**Estimating A.** The main observation here is to realize that since  $|I| \leq |J|$  and  $I \cap (\Omega_J)^c \neq \emptyset$ , one must in particular have  $I_n \cap 3J = \emptyset$ . This allows one to estimate  $A$  by

$$\sum_{|I_n| \leq |J|} \left( 1 + \frac{\text{dist}(I_n, J)}{|I_n|} \right)^{-10} \int_{\mathbf{R}} |b_J(y)| dy \lesssim \lambda |J| \sum_{|I_n| \leq |J|} \left( 1 + \frac{\text{dist}(I_n, J)}{|I_n|} \right)^{-10} \lesssim \lambda |J|,$$

as required by (37)

**Estimating B.** This time, one has to take into account that fact that

$$\int_{\mathbf{R}} b_J(y) dy = 0. \quad (38)$$

As before, one can estimate  $B$  by

$$\sum_{|I_n| > |J|} |\langle b_J, \Phi_{I_n}^\infty \rangle|$$

where this time  $\Phi_{I_n}^\infty := |I_n|^{1/2} \Phi_{I_n}$  is an  $L^\infty$  normalized bump. In order to emphasize that the dependence of  $\mathbf{n}$  is irrelevant now, we rewrite the above expression as

$$\sum_{|K| > |J|} |\langle b_J, \Phi_K^\infty \rangle|$$

where the sum is over dyadic intervals  $K$ .

Fix  $K$  such that  $|K| > |J|$  and observe that

$$|\langle b_J, \Phi_K^\infty \rangle| = \left| \int_{\mathbf{R}} b_J(z) \overline{\Phi_K^\infty}(z) dz \right| = \left| \int_J b_J(z) (\overline{\Phi_K^\infty}(z) - \overline{\Phi_K^\infty}(c_J)) dz \right|$$

where  $c_J$  denotes the midpoint of the interval  $J$ .

Then, observe that for  $z \in J$ , one has

$$|\overline{\Phi_K^\infty}(z) - \overline{\Phi_K^\infty}(c_J)| \lesssim |J| \frac{1}{|K|} \left( 1 + \frac{\text{dist}(K, J)}{|K|} \right)^{-10}$$

and so the previous term becomes smaller than

$$|J| \frac{1}{|K|} \left( 1 + \frac{\text{dist}(K, J)}{|K|} \right)^{-10} \int_J |b_J(y)| dy \lesssim |J| \frac{1}{|K|} \left( 1 + \frac{\text{dist}(K, J)}{|K|} \right)^{-10} \lambda |J|.$$

Finally, the corresponding (37) follows from the straightforward observation that

$$\sum_{|K| > |J|} \frac{|J|}{|K|} \left( 1 + \frac{\text{dist}(K, J)}{|K|} \right)^{-10} \lesssim 1.$$

By interpolating between  $L^2$  and weak- $L^1$  we obtain the theorem for any  $1 < p \leq 2$ . To prove the rest of the estimates we proceed as usual, by duality. Fix then  $2 < p < \infty$ . By using Khinchin inequality, one can write

$$\begin{aligned} \|S^n f\|_p^p &= \int_{\mathbf{R}} \left( \sum_I \frac{|\langle f, \Phi_{I_n} \rangle|^2}{|I|} \chi_I(x) \right)^{p/2} dx \lesssim \\ &\int_{\mathbf{R}} \int_0^1 \left| \sum_I r_I(t) \langle f, \Phi_{I_n} \rangle h_I(x) \right|^p dx dt = \int_0^1 \left\| \sum_I r_I(t) \langle f, \Phi_{I_n} \rangle h_I \right\|_p^p dt, \end{aligned} \quad (39)$$

where  $(r_I)_I$  are the Rademacher functions and  $(h_I)_I$  the  $L^2$ -normalized Haar functions.

Fix now  $t \in [0, 1]$  and consider the linear operator

$$f \rightarrow \sum_I r_I(t) \langle f, \Phi_{I_n} \rangle h_I.$$

Using the fact that  $S^n$  and the Littlewood-Paley square function associated to  $(h_I)_I$  are bounded below  $L^2$ , an argument identical to the one used to prove Theorem 2.3 shows that the above operator is also bounded below  $L^2$  and by duality, above  $L^2$  as well, with bounds independent of  $t$  which grow logarithmically in  $\langle n \rangle$ . Using this fact in (39), completes the proof of the theorem. ■

## 6. GENERALIZATIONS

Let us first observe that the first commutator  $C_1 f$  can also be written as

$$C_1 f(x) = p.v. \int_{\mathbf{R}} \left( \frac{\Delta_t}{t} A(x) \right) f(x+t) \frac{dt}{t} \quad (40)$$

where  $\Delta_t$  is the *finite difference* operator at scale  $t$  given by

$$\Delta_t g(x) := g(x+t) - g(x).$$

There is a very simple way to motivate the introduction of this operator. Start with the Leibnitz rule identity

$$(Af)' = A'f + Af'$$

and solve for  $A'f$  to obtain

$$A'f = (Af)' - Af' = D(Af) - ADf = [D, A]f$$

where  $D$  is the operator of taking one derivative and  $A$  is viewed now as the operator of multiplication with the function  $A(x)$ . In particular, assuming that  $A' \in L^\infty$ , the commutator  $[D, A]$  maps  $L^p$  into itself boundedly, for every  $1 < p < \infty$ . *Does this property hold for the operator  $[[D], A]$  as well ?* one might ask. A straightforward calculation shows that  $[[D], A]$  is precisely the first commutator of Calderón.

Given these, it is of course natural to ask *what can be said about the double commutator  $[[D], [[D], A]]$  ?*

A direct calculation shows that the expression  $[[D], [[D], A]](f)(x)$  is equal to

$$p.v. \int_{\mathbf{R}^2} \left( \frac{\Delta_t}{t} \circ \frac{\Delta_s}{s} A(x) \right) f(x+t+s) \frac{dt}{t} \frac{ds}{s} \quad (41)$$

a formula that can be naturally seen as a bilinear operator, this time depending on  $f$  and  $A''$ . Its symbol can be again calculated easily and it is given by

$$\left( \int_0^1 \text{sgn}(\xi + \alpha \xi_1) d\alpha \right)^2$$

which is precisely the square of the symbol of the first commutator of Calderón.

**Theorem 6.1.** *Let  $a \neq 0$  and  $b \neq 0$  and consider the expression*

$$p.v. \int_{\mathbf{R}^2} \left( \frac{\Delta_{at}}{t} \circ \frac{\Delta_{bs}}{s} A(x) \right) f(x+t+s) \frac{dt}{t} \frac{ds}{s}.$$

*Viewed as a bilinear operator in  $f$  and  $A''$ , it extends naturally as a bounded operator from  $L^p \times L^q$  into  $L^r$  for every  $1 < p, q \leq \infty$  with  $1/p + 1/q = 1/r$  and  $1/2 < r < \infty$ .*

To prove this theorem, one applies the same method described earlier for the first commutator. One just has to observe that the symbol of this operator is given by

$$\left( \int_0^1 \operatorname{sgn}(\xi + \alpha a \xi_1) d\alpha \right) \left( \int_0^1 \operatorname{sgn}(\xi + \alpha b \xi_1) d\alpha \right)$$

and after that to realize that *each factor* satisfies the same desired *quadratic estimates*. So this time one needs to decompose each factor as a double Fourier series as we did before. The fact that one can go all the way down to  $1/2$  with the estimates, is a simple consequence of the statement that series of type

$$\sum_{n_1, n_2 \in \mathcal{Z}} |C(n_1, n_2)|^r \log < n_1 > \log < n_2 >$$

are always convergent as long as the constants  $C(n_1, n_2)$  decay at least quadratically in  $n_1$  and  $n_2$  and  $r > 1/2$ . The details are straightforward and are left to the reader. And clearly, one can generalize the above theorem even further, in the most obvious way. We will come back to this in the second paper of the sequel.

Another generalization we have in mind comes from the following identity

$$A'B' = (AB)'' - (BA')' - (AB')' + A'B'. \quad (42)$$

As a consequence of it, the right hand side of (42) satisfies Hölder estimates of type

$$\|(AB)'' - (BA')' - (AB')' + A'B'\|_r \lesssim \|A'\|_p \|B'\|_q$$

for indices  $p, q, r$  as before. *Does this inequality continue to hold if one replaces every derivative  $D$  by its modulus  $|D|$ ?* As before, a direct calculation shows that the new expression

$$|D|^2(AB) - |D|(B|D|A) - |D|(A|D|B) + (|D|A)(|D|B)$$

can be rewritten as

$$p.v. \int_{\mathbf{R}^2} \left( \frac{\Delta_t}{t} A(x+s) \right) \left( \frac{\Delta_s}{s} B(x+t) \right) \frac{dt}{t} \frac{ds}{s}. \quad (43)$$

The right way to look at this formula is to view it as a bilinear operator in  $A'$  and  $B'$ . Its symbol can be calculated quite easily and it is given by

$$\left( \int_0^1 \operatorname{sgn}(\xi_1 + \alpha \xi_2) d\alpha \right) \left( \int_0^1 \operatorname{sgn}(\xi_2 + \beta \xi_1) d\beta \right) \quad (44)$$

which is a symmetric function in the variables  $\xi_1$  and  $\xi_2$ . Because of this symmetry we like to call expressions such as the ones in (43) *circular commutators*. We will return to them in the second paper of the sequel.

**Theorem 6.2.** *Let  $a \neq 0$  and  $b \neq 0$  and consider the expression*

$$p.v. \int_{\mathbf{R}^2} \left( \frac{\Delta_{at}}{t} A(x+s) \right) \left( \frac{\Delta_{bs}}{s} B(x+t) \right) \frac{dt}{t} \frac{ds}{s}.$$

*Viewed as a bilinear operator in  $A'$  and  $B'$ , it extends naturally as a bounded operator from  $L^p \times L^q$  into  $L^r$  for every  $1 < p, q \leq \infty$  with  $1/p + 1/q = 1/r$  and  $1/2 < r < \infty$ .*

The proof uses the same method, since it is not difficult to see that the symbols of such bilinear operators are again products of symbols of the first commutator kind and they each satisfy the same *quadratic estimates*.

#### REFERENCES

- [1] Calderón A., *Commutators of singular integral operators*, Proc. Nat. Acad. Sci. U.S.A., vol. 53, 1092-1099, [1965].
- [2] Calderón A., *Commutators, singular integrals on Lipschitz curves and applications*, Proc. Int. Congress of Math., Helsinki, 1978. Academia Scientiarum Fennica, Helsinki, 85-96, [1980].
- [3] Calderón C., *On commutators of singular integrals*, Studia Math., vol. 53, 139-174, [1975].
- [4] Christ M., Journé J.L., *Polynomial growth estimates for multilinear singular integral operators*, Acta Math., vol. 159, 51-80, [1987].
- [5] Cohen J., Gosselin J., *A BMO estimate for multilinear singular integrals*, Illinois J. Math., vol. 30, 445-465, [1986].
- [6] Coifman R. R., Jones P., Semmes S., *Two elementary proofs of the  $L^2$ -boundedness of the Cauchy integral on Lipschitz curves*, J. Amer. Math. Soc., vol. 2, 553-564, [1989].
- [7] Coifman R. R., Meyer, Y., *On commutators of singular integrals and bilinear singular integrals*, Trans. Amer. Math. Soc., 315-331, [1975].
- [8] Coifman R. R., Meyer, Y., *Wavelets. Calderón-Zygmund and multilinear operators*, Cambridge Studies in advanced mathematics, [1997].
- [9] Coifman R., McIntosh, A. and Meyer, Y., *L'intégral de Cauchy définit un opérateur borné sur les courbes Lipschitziennes*, Ann of Math., vol. 116, 361-387, [1982].
- [10] David G, Journé J.L., *A boundedness criterion for generalized Calderón-Zygmund operators*, Ann. of Math., vol. 120, 371-397, [1984].
- [11] David G, Journé J.L., Semmes S., *Opérateurs Calderón-Zygmund, fonctions para-accrétive et interpolation*, Revista Mat. Iberoamericana, vol. 1, 1-56, [1985].
- [12] Fefferman C., *Recent progresss in classical Fourier analysis*, Proceedings of the ICM, Vancouver, 1974, vol. 1, 95-118, Canad. Math. Congress, Montreal, QE, [1975].
- [13] Grafakos L., Li X., *Uniform estimates for the bilinear Hilbert transform I*, Ann. of Math. vol. 159, 889-933, [2004].
- [14] Kenig C., *Weighted  $H^p$  spaces on Lipschitz domains*, Amer. J. Math., vol. 102, 129-163, [1980].
- [15] Lacey M., Thiele C.,  *$L^p$  estimates for the bilinear Hilbert transform for  $2 < p < \infty$* , Ann. of Math., vol. 146, 693-724, [1997].
- [16] Lacey M., Thiele C., *On Calderón's conjecture*, Ann. of Math., vol. 149, 475-496, [1999].
- [17] Li X., *Uniform estimates for the bilinear Hilbert transform II*, Revista Mat. Iberoamericana, vol. 22, 1069-1126, [2006].
- [18] Mitrea M., *Clifford Wavelets, Singular integrals and Hardy spaces*, LNM, vol. 1575, Springer Verlag, [1994].
- [19] Murai T., *A real variable method for the Cauchy transform and analytic capacity*, L.N.M. vol 1307, Springer-Verlag, Berlin, [1984].
- [20] Muscalu C., *Paraproducts with flag singularities I. A case study*, Revista Mat. Iberoamericana, vol. 23, 705-742, [2007].

- [21] Muscalu C., Pipher J., Tao T., Thiele C., *Bi-parameter paraproducts*, Acta Math., vol. 193, 269-296, [2004].
- [22] Muscalu C., Pipher J., Tao T., Thiele C., *Multi-parameter paraproducts*, Revista Mat. Iberoamericana, vol 22, 963-976, [2006].
- [23] Muscalu C., Tao T., Thiele C., *Multi-linear operators given by singular multipliers*, J.Amer.Math.Soc., vol. 15, pages 469-496, [2001].
- [24] Palsson, E. *Ph.D. Thesis*, Cornell University, [2011].
- [25] Stein E., *Harmonic analysis: real-variable methods, orthogonality and oscillatory integrals*, Princeton University Press, [1993].
- [26] Thiele C., *A uniform estimate*, Ann. of Math., vol. 156, 519-563, [2002].
- [27] Verdera J.,  *$L^2$  boundedness of the Cauchy integral and Menger curvature*, Contemporary Mathematics, vol. 277, 139-158, [2001].

DEPARTMENT OF MATHEMATICS, CORNELL UNIVERSITY, ITHACA, NY 14853  
E-mail address: camil@math.cornell.edu